DOCUMENT RESUME

ED 052 236 TM 000 640

AUTHOR Gephart, William J.; Bartos, Bruce B.
TITLE Profiling Instructional Package.
INSTITUTION Phi Delta Kappa, Bloomington, Ind.
REPORT NO Occas-Pap-7
PUB DATE Aug 69
NOTE 52p.

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Data Analysis, Data Collection, Data Processing, Educational Research, Evaluation, *Flow Charts, Hypothesis Testing, *Instructional Materials, Logical Thinking, *Profile Evaluation, *Research Methodology, Statistical Analysis

ABSTRACT

The necessity of evaluating research methodology is discussed, and a profiling approach is proposed. Explanatory statements about the research process, directions for doing the actual profiling of the research report, a research profiling flow chart, a research profile sheet, and two additional graphic aids to understanding are included. The procedure outlined in the Research Profiling Flow Chart first examines whether the research is a test of a hypothesis, or an answer to an empirical question; it then focuses on the generation of data. The final facet of the research process is data analysis. For further information on profiling, see Occasional Paper #3: Profiling in Educational Research (TM 000 643). (AG)

PHI DELTA KAPPA

RESEARCH SERVICE CENTER

OCCASIONAL PAPER

#7 PROFILING INSTRUCTIONAL PACKAGE

by

William J. Gephart
Director of Research Services
Phi Delta Kappa

Bruce B. Bartos
Research Assistant
Phi Delta Kappa

August, 1969

INTRODUCTION

Why evaluate the methodology of educational research? Why not accept the facts stated in published research reports? Two things provide the bases for the answers to those questions. The first is found in the nature of the research activity itself. The second relates to the adequacy of already existing techniques employed by researchers. The latter point is quite straightforward; our techniques are far from fool-proof! The former is a little more complex. For an investigation, the researcher designs a situation which *approximates* the reality of the problem he is studying. Since there are some unknowns in any situation, these approximations will not coincide with reality. For these two reasons, the user of the research results is cautioned against a blind acceptance of them. Unless all the details of the research activity are understood, the findings *must* be considered with reservations.

Research activities are multifaceted. The procedure outlined in the Research Profiling Flow Chart takes into account those facets. It examines first whether the research is a test of a hypothesis or an answer to an empirical question. Although a test of the hypothesis inherently has more levels of soundness both should present logical arguments. The flow chart traces this logic. Next, the flow chart focuses on the generation of data. Three elements are explicated here: the representativeness of the units studied, the treatments or experiences common to those units, and finally, the manner in which the phenomena have been measured. The final facet of the research process is data analysis. Each of these segments will be explained in detail, but first, a brief statement is needed regarding the basic approach proposed, that is, profiling.

PROFILING: AN APPROACH TO THE EVALUATION OF RESEARCH REPORTS

These materials are designed to assist the individual in PROFILING, a form of evaluation of *educational research. They* are, in effect, a set of instructional materials including three interlocking components. The first component consists of EXPLANATORY STATEMENTS about the research process, in general. The second component consists of DIRECTIONS for doing the actual profiling of the research report. A RESEARCH PROFILING FLOW CHART, a RESEARCH PROFILE SHEET, and two additional graphic aids to understanding make up the third component. The reader should imagine himself as a PROFILER in a situation requiring *justifiable decisions about* accepting or rejecting the conclusions stated in a given research report.

To aid the Profiler several statements need to be made about *the* format of this presentation. The first and second components are differentiated by their print layout. More specifically, the first component (explanatory statements) uses wide margins and this type face. In contrast, the directions component is indented and in a different type face. The charts and figures of the third component are located at the end of this packet, on fold-out pages. This enables the figure to be viewed while the several related pages of text are being studied. As concepts and keywords basic to the process of profiling are presented, they are *HIGHLIGHTED* by *capitalization.*

At this point, Figure 1 on Page 39should be folded out for viewing while reading the next section. It presents a block diagram on the FACETS OF THE RESEARCH PROCESS, which these materials cover. The PROFILE in the bottom-left corner is a distillation of the final product of this package - the Research Profile Sheet. The Profile will also be found in the upper-right hand corner of the profile sheet.

As already indicated, research methodology has many facets and *it* involves *an inherent logical* argument, the selection of subjects to be studied, structuring of experiences for those subjects, measurement related to those experiences, and analysis of the collected data. Most reports of research present sound procedures in some of these facets and weak procedures in others, a condition that precludes a single statement, "This research is good/bad!"

There is a second reason for not making a single statement about the soundness of the conclusions. The use of research findings is complicated by variation in the need for information in professional decisions. There are times when one must use conclusions which are at best only tentative. There are other times in which the need for information requires almost a *guarantee* of truth. For example, if the need for knowledge in an area is great and the amount of risk to personal safety is relatively low, conclusions can be accepted and operated on despite weaknesses in the procedures used to generate those conclusions (Example, a study of a process to produce cheaper textbooks). On the other hand, if there is a possibility of personal injury, conclusions based on weak procedures cannot be tolerated. Consider for a moment the work on the development of a typical virile strain vaccine against rabies. Preliminary findings can give us sufficient information to work with the vaccine on animals. But until the research methodology is firmly established, we are morally restrained from conducting research which involves humans. Research methods in the social sciences seldom invoke such drama; however, the same concerns exist.

There is one more reason why a completed peice of research and its conclusions cannot be labeled "good" or "bad". That reason deals with the unpredictable future uses of a research finding. A research effort is completed and stored in the professional literature. However, it may eventually be put to use in any number of ways by different professionals, ways which preclude the possibility of determining once and for all the value of each particular effort.

Regardless of the knowledge needs or professional circumstances, a given conclusion ought not to be accepted, held tentatively, or rejected without careful evaluation of the research procedures used to generate it. Once the research user under-stands the strength of the procedures and the various facets of the completed research, he is in a better position to use the conclusions of that research in professional decisions. It is asserted that the profiling procedure in this paper will facilitate the labeling of the methodology of completed research reports, and thus aid the research utilization process.

FACETS OF THE RESEARCH PROCESS ELEMENTS OF PROFILING

The conducting of an empirical study requires several steps. Those steps are the elements upon which the profiling activity focuses. They include: (1) the structuring of a logical argument, (2) the generation of data, and (3) the statistical analysis of data. All three items are involved in investigations which test hypotheses, while only (2) and (3) are used in studies which attempt to answer empirical questions.

Now refold the FACETS OF THE RESEARCH PROCESS page fold-out and continue reading.

Logic

Before detailing these elements it seems important to define the meaning of the terms, TEST OF A HYPOTHESIS and ANSWER TO AN EMPIRICAL QUESTION, as they are used throughout this document.

Two mutually exclusive categories of questions exist; empirical and non-empirical. The word "empirical" connotes a direct observation. An EMPIRICAL QUESTION, then, is one for which there is possibly a direct observation to determine its answer. Non-empirical questions require a combination of direct observation and inferences about the observation and related concepts. A couple of quick and simple examples will solidify these statements. How many test items on the arithmetic section of the California Achievement Test will be answered correctly by a specific child? To answer that question, the test is administered, scored, and the items marked correctly are counted. The number correct *is* the direct observation. Consider the same situation but a question which cannot be labeled "empirical". Does a specific child understand arithmetic? The answer to that question *is not* directly observable. Understanding is a *mental* activity not visible or countable in any direct manner. An answer can be generated; however, it requires the selection of some directly observable activity is a logical indicator of understanding. In this case we probably would look at *both* the number of items answered correctly *and* the grade level of the child.

A HYPOTHESIS in the context of research is a tentative statement asserting a relationship between two or more things (Kerlinger, 1964). Some example hypotheses will help to make the point:

1. Teachers provide a different instructional treatment for boys than they do for girls.
2. Instructional method A is better than instructional method B.
3. Numerous short periods of memorization time are better than one concentrated period.

The analysis of the elements of these hypotheses requires the understanding of the term VARIABLE. In this presentation variable refers to anything that can exist in different quantities or qualities.

In the first hypothesis the two things. the two variables. are instructional treatment and sex. It asserts that there is a different quality of instructional treatment for the different qualities of sex. A TEST OF that HYPOTHESIS is an attempt to determine whether or not this is in fact true.

The second hypothesis also contains two variables. One of them is explicitly stated: the other implied. The explicitly stated variable is "instructional methods". Two qualities. method A and method B. of that variable are delineated in the hypothesis. The impled variable is the "better" output from that instruction. It might take one of several forms. including achievement. speed. utility, etc. In the report of a test of that hypothesis. the exact nature of the implied variable would have to be delineated *through the procedures used by the investigator.*

In most research activities it is impossible to make a direct observation to determine the truth of a hypothesis. Consider the third hypothesis above: short periods of practice are better than one extended period for memorizing a passage. Memorization exists within the mind of the individual. We cannot tell if memorization has occurred without calling for some other more observable activity. The activity most typically used is the recall of the memorized material. Thus, the researcher reasons that if the hypothesis is true, certain things should be directly observable. In this case, better recall of the passage will be seen on the part of the students who use the repeated short periods of practice than on the part of the students who use the one extended practice period.

The third hypothesis also relates implicit and explicit variables. The explicit variable is the approach to memorization, either repeated short periods or one extended period. The implicit variable is the quality of the results, that is, the degree to which the individual memorized the material. Again, a test of the hypothesis would have to make explicit the nature of that second variable. This analysis should make clear that hypotheses are statements about variables AND relationships. Sometimes these are made explicit, other times implicit.

An hypothesis is an *assertion*, the truth of which must be determined. One says, in effect, "Here is a statement I believe to be true. I will now engage in certain activities to establish the soundness of that belief." Those activities constitute tests of the hypothesis. And until they are conducted, the truth value of the assertion is minimal.

The Structuring of a Logical Argument

A LOGICAL ARGUMENT is inherent in a test of a hypothesis. It is described by the mathematical logician, George Polya (1954) as consisting of a major premise, one or more minor premises, and a conclusion.

The MAJOR PREMISE is typically a statement which asserts

"If the hypothesis is a true statement, then specified events will be observed as indicators of that truth." An example of a major premise can be seen in relation to the first hypothesis in the list above. That hypothesis says that teachers give different instructional treatments to boys than they do to girls. As indicators of the truth of that statement, a researcher (McNeil, 1964) reasoned that boys would be nominated more often than girls as the persons to whom the teacher directed certain kinds of actions. This major premise could be stated as,

> If the hypothesis (teachers provide a different instructional treatment for boy than they do for girls) is a true statement, then systematic differences by sex will be seen when children are asked to name the students who receive specified teacher treatments.

The first part of this major premise is an indication that a researcher is seeking evidence on which to judge the truth of a hypothesis. The second part, that is, the actions which are to be observed and used as indicators are called the CONSEQUENTS.

Two kinds of MINOR PREMISES have been evolved from Polya's work (Raths, 1969). The first deals with the observation of the consequents. Where they or were they not seen? The exact nature of this minor premise in a given study is determined *after*: (1) a situation is structured in which the consequents should occur; (2) that situation is observed; and (3) data from those observations have been analyzed. In the example referred to above, significant differences were observed. The minor premise in this case was, "There is a systematic sex differentiation in the nominations." Note: This premise *does not say* that the hypothesis is true. Rather, it focuses on the observations. There was a difference seen in the nominations.

The second category of minor premises is necessary because of MULTIPLE CAUSATION principle in research in the social sciences. Any event which we observe is likely to be shaped by many factors. Consider the event: obtaining significant differences by sex of the students named as recipients of specified teacher treatments. Factors which contribute to that event range from systematic differences in the groups of boys and girls in terms of age, maturity, and intelligence, to using teacher-student interactions that *must* be differentiated by sex. If one asks to whom does the teacher say, "Zip up your fly!", she ain't talkin' to girls.

Because of this MULTIPLE CAUSATION PRINCIPLE, a test of any hypothesis must concern itself, in part, with all of the prossible causes for an observation. In conducting such a test, an investigator establishes circumstances in which observations are made. Those observations are the possible result of many factors. One of those factors is or should be the relationship expressed in the hypothesis. The other factors are RIVAL HYPOTHESES (alternate explanations for the event). When he structured his study, the researcher ought to have done things which ruled out those rival explanations. The second category of minor premises covers the extent to which those rival hypotheses *are* ruled out.

Completed research could be divided into three categories with respect to the second minor premise. The first would include those efforts in which *no* rival explanations are apparent. In the second category rival explanations *may* exist. In the final, rival hypotheses are difinitely involved.

The last element of a logical argument is the CONCULSION. Its form in a given study is dependent upon the nature of the *two* minor premises. This is because the first of these presents information as to whether or not the observation was made, while the second indicates whether rival hypotheses are present. If the consequences predicted are observed, support for the truth of the hypothesis is presented. If the observation is not made, support cannot be claimed. (Note: Failure to make the predicted observation does not automatically mean rejection of the hypothesis.) The second minor premise determines the strength of the conclusion. If rival explanations are known, very weak support of the truth of the hypothesis has been developed. If there is the possibility but not the probability of rival explanations, tentative support is generated and the hypothesis is at least credible. And finally, if no rival explanations or rival hypotheses are conceivable, the truth of the original hypothesis is very much more credible.

At this point, the logic section of the Research Profiling Flow Chart (page 41) and the Research Profile Sheet (page 47) should be folded out. The flow chart consists of RECTANGLES, DIAMONDS, CIRCLES, and PRINTOUT-SYMBOLS. The rectangles indicate activities engaged in; diamonds represent questions about these activities; and circles delineate specific alternative answers to these questions. The pattern of connections between these symbols leads an individual through the chart. At several points the user can find a printout-symbol - a four-sided figure with one curved side. This indicates a LABELING that is to take place on the Research Profile Sheet. When one of these figures is encountered, the Profiler marks the profile sheet as directed by the statement within the figure, *and proceeds to the next section on both the flow chart and these written instructions.* A trapezoid and an elongated hexagon have special purposes which are obvious upon inspection.

Work with the Research Profiling Flow Chart starts in the upper left corner. The trapezoidal input box indicates that the Profiler has a report from the professional literature. His first instruction is to examine that report for a presentation of DATA. Data in this instance mean numbers, scores, or frequencies of observations that have been made during the research. The first question raised simply asks whether or not data are presented. If they are not, the analysis activity is stopped and the appropriate label is checked on the profiling sheet. The report is either not a research effort *or* it is an incomplete part of the research process.

If data are presented, however, the path leads through connector number 1, and the Logic sector of the chart. The discrimination between a test of a hypothesis and an answer to an empirical question is undertaken. To make the discrimination the individual is asked first to examine the report for the existence of an hypothesis. The definition of an hypothesis was given on page 5; it is repeated in the activity box as two or more variables and a predicted relationship between them. The next question faced in the evaluation process is, "Is a hypothesis presented?" Let us consider the results of a negative answer (exit Q) before we take up the positive ones.

Examine the text of the report to see if specific questions are posed and answered. If they are not, stop all further work with the chart and mark the "Stop" space at the top of the Research Profile Sheet. If, however, questions *are* posed and answered, two possibilities still exist. As the Profiler will remember from the material presented *earlier*, questions can be either empirical or non-empirical. The next decision in the flow chart deals with that discrimination. "Are the questions answerable by direct observation?" If the answer is "Yes", a label is required. The labeling of a report as an "Answer to an empirical question" refers to the profile sheet again and asks for a check at space Q under Logic. Once that check is made, the Profiler proceeds as directed by the arrow to connector number 2 on the flow chart. If the answer is "No", one further decision needs to be made.

Consider again the examples used to distinguish between empirical and non-empirical questions in the earlier discussion. The empirical question was, "How many test items on the arithmetic section of the California Achievement Test will be answered correctly by a specific child?" The non-empirical question was, "Does a specific child understand arithmetic?" In this latter case, there is a logical combination of referents that is obserable - score and grade level.

On the flow chart, a "Yes" in response to "Is a logical referent observed to answer the question?" also causes us to mark space Q on the Research Profile Sheet and to move on to connector number 2. A "No" indicates a "Stop" label and a termination of further activity with this report.

Return to connector number 1. If the question "Is a hypothesis presented?" is answerable with "Explicitly" through specific statements or "Implicity" in the content of the work, both are treated alike and exit H is taken.

At this point the report has been analyzed either as an answer to an empirical question or as a test of a hypothesis. If it is the latter, the logical argument underlying that test must be delineated. As shown on the flow chart, this is initiated by identifying and stating the hypothesis and the consequents which are to be observed. In effect, the Profiler is asked to determine the major premise in the logical argument. That identification task will require careful reading of the research report. The Profiler should note either the explicit or the implicit hypothesis being tested, and the observations which the researcher is using as evidence that his hypothesis is true. Here a concern is introduced about the major premise that was not alluded to in the earlier discussions. There *must* be a logical and acceptable connection between the hypothesis and the items to be observed. And absurd example may help to make the point. A test of the hypothesis that method A is better than instructional method B *could not* be made by observing the average annual wind velocity in a given community. There is no *logical connection* between the hypothesis and the specified consequents in this case. Nothing about wind velocity attests to the difference in quality of the two instructional methods.

Now return to the decision box on the flow chart which asks "Are the consequents logical if the hypothesis is a true statement?" If a negative response is obtained, the Profiler is directed to "Stop" and label since it is useless to consider further an invalid test of a hypothesis. A positive response leads to the activity box which requires the location of the section of the research report dealing with the observation of those consequences. In this case the focus is on the *first minor premise* in the logical argument. "Were the consequents observed?" For the answer to that question the Profiler should check the data analysis section of the report. Look for statistics which indicate significance in the observations. Three possible answers, two of which lead to the same conclusion, are shown in the Flow Chart. Those two are: the report either does not say that these events were observed, or it says in fact they were not observed. In either of these cases a "No conclusion" label is required. That is, no conclusion can be made at this point regarding the truth of the hypothesis. Given that label and interest in pursuing the details of the study further, the individual goes to connector number 2.

If those consequents were in fact observed (the third possible answer), the *second minor premise* needs to be examined. An earlier discussion said that this premise deals with rival explanations for the observation. The flow chart refers to these as RIVAL HYPOTHESES. The Profiler examines the entire report to identify the existence of possible alternate explanations for the events observed.

There are two general sources of rival explanations: variables inherent in the problem area being studied; and variables related to the research process itself. As an example of the first general source, consider again the hypothesis: Instructional method A is better than instructional method B. Suppose two groups have been established, one taught by method A, one by method B. A common achievement test was used at the end of the instructional program. Significant differences were seen which indicate greater achievement on the part of the students who used instructional method A. If the group of students who were taught by method A were more intelligent than the group that experienced method B, a rival explanation (rival to the hypothesis being tested) would be provided. The same situation would exist if the group of students receiving method A were systematically older or if that teacher were a noticably better teacher. These and other variables related to the treatment are *possible* rival explanations for the observation that the group which received method A outscored the method B group.

The second source of rival explanations is the research process itself. The observations can sometimes be explained as a direct result of the *procedures* followed by the researcher. Consider an investigation in which the group taught by method A received the treatment at a time when the students were fresh and alert, and the group receiving method B was scheduled for a period when they were tired and drowsy. This time- sequence factor could explain the observed differences.

Another example of a research-process-caused rival explanation is *cueing*. Pretests can alert students to the content of instructional material in the treatment to follow. If the pretest is designed so that it has a *bias* toward method A, this bias is a rival explanation for the observation. An excellent list and discussion of the rival explanations that stem from the research process itself has been presented by Campbell and Stanley in the HANDBOOK OF RESEARCH ON TEACHING (1963).

After the profiler has examined the report for possible rival hypotheses or rival explanations for the observed results, he asks himself the question, "Are rival hypotheses (1) known to be present, (2) possibly in the study, or (3) not identifiable?" The answer to this question leads again to a label under the Logic section of the Research Profile Sheet. It should be noted that this question, and its answer, establishes the relative strength of the conclusions that can be formed from the research findings. That strength ranges: from the truth of the hypothesis is verified (always short of absolute proof), to the truth of the hypothesis is credible, to the other extreme, the truth of the hypothesis is questionable. Once this label is attached as directed by the profiling chart, the Profiler moves to Phase 2.

Note: Only one label should be checked for each study. If you have checked more than one label on the Profiling Sheet, you should go back to the start of the Logic section, re-read the general statements, and follow the directions again.

The LOGIC section of the profile sheet has now been completed, indicating that the logical bases for the study have been evaluated. Refold the LOGIC sheet and fold out the Data Quality sheet (page 40).

DATA QUALITY

As indicated earlier, a researcher identifies a question to be answered or a hypothesis to be tested. To achieve either of these, he structures a situation in which he generates or accumulates bits of information called DATA. These bits of information might exist in the form of numbers or in the form of repeated verbal statements. In the first case these are more likely called "scores," in the second the term "frequencies" is appropriate. The data for any given study are shaped by the procedures followed by the investigator.

The Generation of Data

The second major facet in the research process, the generation of data, contains three elements. If variation occurs in any of these three, a different set of data is generated. REPRESENTATIVENESS is the first of these elements. Consider an investigation of the effects of test anxiety on student achievement. A test of test anxiety is administered to graduate students, and the high and low 25 per cent are selected as subjects. The results from the achievement test will yield a particular set of data. On the other hand, if the study uses a randomly selected group of high school seniors, an entirely different set will be produced. And neither group of subjects is representative of students in general.

The second way to cause variation in the data is the TREATMENT or experiences of the subjects. Again the test anxiety area provides an example. One set of data could be generated by a treatment in which the subjects are given an achievement test that was constructed for students at a much higher level of education. Consider the same group of students but a slightly different treatment. In this case, a test at the appropriate educational level is administered repeatedly. Each day that it is given the test is described as an exact replica of one which will be given to them in the not-too- distant future. That future test will determine whether or not they are allowed the educational program of their choice. The focus is still the effects of test anxiety, but the shape of the data will be somewhat different than in the first case.

The third shaping aspect in data generation is MEASUREMENT. If the effects of a specified treatment on a specified group are measured by a paper and pencil test such as the Mandler-Sarason Test Anxiety Questionnaire, one set of data will be generated. If, however, the arm rests of the chairs in the classroom were wired for galvanic skin-response measurements, a quite different set of data would be obtained.

These three aspects of data generation, as related to data quality, can be displayed graphically with the cube (Gephart & Ingle, 1969) displayed on the Data Quality Cube fold out. The height of the cube stands for representativeness or sample quality (OA). This is the degree to which the units studied characterize a specified population. Treatment or experiences is represented on the width of the cube (OC). And measurement quality is displayed by its depth (OG).

The range of quality on the representativeness dimension is from a high (point A on the cube) at which perfect representation of a specified population is assured, to a low (point O) at which some unspecified units were studied. Quality points between these extremes will be discussed later.

The treatment scale of quality has a similar pair of dimensional extremes. At point C would be those studies which define the treatment in terms of its character, sequence, and duration. The other extreme (point O) would indicate those cases where the units studied have some common but undefinable set of experiences. In such a case the researcher is unable to state definitively what actually happened to the subjects.

The third dimension of the cube, measurement, has a low (point O) at which some records were kept - but probably for other purposes than the research study in question and in a manner which leaves their validity and reliability unknown or at least in doubt. The other extreme (point G) would be an instance in which the measurement was carried out through perfectly objective, valid, and reliable techniques.

One can hypothetically locate research projects on or in this cube. For example, a project which either used a total specified population or selected a perfectly representative sample of that population would be located at point A. If the content and sequence of all of the treatments employed were completely detailed in the report by the researcher, the project would then be at point B. Finally, if the measuring techniques were perfectly objective, valid, and reliable, the study would rest at point E. Any given study seldom reaches this level of data generation quality.

Rather, it falls somewhere between the extremes, either on one of the faces of the cube or somewhere inside the cube.

Dimensions for the Data Quality Cube

The previous section sketched in the extremes on the data quality cube. The evaluation of a specific piece of research requires the identification of points between these extremes. Scales for the three factors in data generation, representativeness, treatment, and measurement, will be detailed in the discussion which follows. After that, the profiling flow chart areas dealing with these scales will be discussed.

The representativeness scale handles the dual question, who was studied, and whom do those units studied represent? The research project is undertaken for the purpose of reaching a conclusion. Conclusions do not float in a vacuum. They are related to a particular time and setting. A given conclusion, then, has at least two aspect: it is about some thing; and it is applicable in some setting. In this respect, two terms are common in the language of researchers. They are "population" and "sample." A POPULATION is a total set of persons or things included in a discrete group which can be described on a specific set of variables. A SAMPLE is a fraction of the population, and can be described by the same set of variables. A sample may or may not be an accurate representation of the population of interest. To be an accurate representation, a sample must display proportionally the same distribution as does the population on all the relevant variables. Consider the following example: all of the students in grades 7, 8, 9 enrolled in Washington, D.C. public schools on March 14, 1969 could be a population? they are certainly a discrete group. The set of variables which would describe that population include: (1) occupational status--students; (2) grade of enrollment--7, 8, 9; (3) date--March 14, 1969; (4) location--Washington, D.C. The four variables used in describing this example population are of an inclusion/exclusion nature. To be a member of that population, all four must be satisfied. To be a member of a sample from that population, an individual must also display all four of these characteristics. A sample of that population would be a fraction of the students in the specified grades at the specified time in the specified location.

In order to determine whether that sample is *representative* of the population, the distributions on the four variables above and distributions on additional varibles related to the topic being studied must be examined. That is, on the variable "occupational status," the population would be described as 100 per cent student, 0 per cent for anyother occupational category. On the variable "grade of enrollment," fractions of the population would be distributed at the 7th, 8th, and 9th grade levels. The sample, to be representative, must display the same attributes, that is, 100% of the individuals in the sample must have the occupational status of students. And, the proportion of the sample enrolled in the 7th, 8th, and 9th grades must be identical to that found in the population as a whole. These along with date and location items would be the basic inclusion/exclusion criteria.

The other variables to be considered to determine the representativeness of this sample are outside of the inclusion/exclusion categories. They would include for example, sex and intelligence, as those are other relevant variables on which the population listed above could be described. In a specific research project these two variables may be related to the variables in the hypothesis being tested. Thus, to be a representative sample the individuals to be included in a study would have to display proportionally the same distribution of sex and intelligence as exists within the population. In any given study still other relevant variables amy have to be considered to determine if a representative sample has been selected.

There are several ways in which subjects or units are selected by researchers. These range from the simple use of available units to the examination of the entire population. Between these extremes are sampling activities involving random selection, purposive sampling, and the solicitation of volunteers. RANDOM SELECTION involves procedures which guarantee that every member of the population has an equal probability of being selected. There are numerous variations in random sampling appropriate for varying sizes of populations and/or varying purposes in investigations. These variations include stratified random sampling, random cluster sampling, and two- or three-stage random sampling. The essence of each of these, however, is that each element in the population has an equal probability of being selected as a part of the sample.

PURPOSIVE SAMPLING involves deliberate decisions and actions on the part of the researcher. After reasoning that a population can be subdivided into specific categories, he searches through the population and selects units in each of those classifications.

At times investigations are undertaken either in settings or on topics which require that the subjects VOLUNTEER for participation. The characteristics of these volunteers cannot be assumed to be generally held by the population to which the investigator may wish to generalize. The very fact that some individuals choose *not* to volunteer while others do represents some differences. Typically, the characteristics which lead to volunteering are not known or explicated in the specific study.

Given the above discussion of the extremes and intermediary points, the REPRESENTATIVENESS SCALE or unit quality has five points.

Those are:

- R_1 = an unidentified group of subjects was studied.
- R_2 = volunteers were studied.
- R_3 = purposive sampling from a specified population established the group studied.
- R_4 = random selection from a specified population established the group studied.
- R_5 = the entire population was studied.

(The symbols R_1, R_2 . . . will be used in connection with these statements in the profiling activity),

As one moves from the top to bottom of this list, the representativeness of units studied improves in quality. These units structure the next section of the research profiling chart.

Before stating this section, remember that as a profiler you are trying to find one statement from the list above that best describes the sample involved in the study being evaluated. That statement will be found when you work through the next flow chart to a symbol that says LABEL. When you reach that point go on to the next section. Close page fold out and open the *DATA QUALITY-REPRESENTATIVENESS* fold out, page 42. Here the use of the Research Profiling Flow Chart continues at connector number 2. The Profiler is directed to identify the population of interest and the sample studied. This information should be stated explicitly in the report. If not, it can sometimes be determined by examining various sections of the report. For example, if the population of interest is not explicitly described in the introduction of the research problem or in the procedures section, it can sometimes be inferred by examining the conclusions made. In or near his conclusions the researcher usually summarizes the central characteristics of his study, including references to the larger group from which the sample was drawn. The reader should realize that a research report *ought* to be more explicit. The proper location for population characteristics is the procedures or data analysis sections.

In this phase, the first question faced by the Profiler is, "Does the report delineate the population to which the generalizations apply?" The term "delineate," in this question, infers that the researcher ought to know, and state, the population boundaries, that is, the inclusion *and* exclusion variables which describe them Furthermore, he ought to have specified the nature of the population on variables which are related to the subject being studied. If the answer to that question is negative, that is, the population of interest is *not* delineated, the Profiler is advised to identify the units studied and answer the question, "Are the studied units described in terms of their distribution on relevant variables?" If this has not been done, the label R_1 is checked on the Research Profile Sheet. Such a project would be an instance in which an unspecified group was observed and in which the reader does not know to whom the findings and conclusions apply. If these units are described, it may be possible to identify the population by extrapolation. The conclusions and findings would then apply to similar groups. An affirmative answer to the question above directs the Profiler to treat the report as though the population were indeed thoroughly specified.

The next activity proposed in evaluating a given piece of research is the identification of how the units studied were selected. A series of questions follow this identification procedure. Each question leads either to a label or to the next question. The first of these is, "Was the entire population studied?" If the answer is "Yes," the label R_5 is checked on the Research Profile Sheet. If the entire population was *not* investigated, the Profiler asks the question, "Was randomization employed to select a sample from the population of interest?" The focus here is on *whether* randomization was employed, not on which randomization technique. It should be understood at this point that this is a random selection of subjects from a specified population, in contrast with random assignment of subjects to specified treatments. If the answer to *this* question is affirmative, the label R_4 is checked. If the answer is negative, still a third question is raised. "Were the units selected through deliberate or purposive procedures, from volunteers, or by some unknown means?" Again, these items lead to labels: Deliberate selection to R_3; Volunteers to R_2; and Unknowns means to R_1. Once these decisions have been reached the Profiler is directed to connector number 3, which leads to an analysis of the treatment characteristics of the study. There is one exception that must be considered first. Does the report have more than one sample? (For example, is there a sample of students AND a sample of teachers?) If so, repeat this entire section for the second sample.

Now refold page 42 - Representativeness.

The treatment dimension on the Data Quality Cube has been described as ranging from the situation in which some unknown treatments were experienced by the subjects, to the other extreme in which the details of the experience are completely known and controlled. Four additional quality levels can be described between these two extremes. The set of six levels makes up the treatment scale. It is readily divisible into two equal groups. One group covers those instances in which the researcher states a theory. The other gourp covers those projects in which a theoretical base for the variables studied is missing (not presented). In this context a THEORY is a formualtion of apparent relationships or principles underlying certain observed phenomena which have been verified to some degree. It consists of the identification of the variables that are involved and/or interacting in a system. A theory should also state what is known about the variables and about the manner in which they interact.

Three categories of variables can be described. The first includes those variables in either the hypothesis being tested or in the empirical questions being answered by the study. The second category (MEDIATING VARIABLES) includes those recognized in the theory as related to, affecting, or interacting with the specific variables being studied. The third category encompasses those variables which, according to existing information, cannot be included in the theory but which *might* have an effect on the variables central to the research effort itself. These are called EXTRANEOUS VARIABLES in this presentation.

Let it be asserted here that theory-based research is of higher quality than atheoretical works. Problems and unknowns in the field of education must exist as components of *some* system. When a hypothesis is stated, at least two variables are made explicit. Many other variables are involved, however. The relationship between those variables stated in the hypothesis and the other associated varibles is an important definer of the system being studied. The same assertion is made about a study which answers an empirical question. Such studies typically investigate the manner in which the subjects distribute along some single variable.

When a theory is stated, quality of a given treatment is dependent upon the degree to which control is asserted over all of the variables. The lowest possible level in this respect would be control over the variables being studied. The next level would include this control plus control over additional variables seen related in the theory. There is still one better level. The state of our knowlege about the field of education is such that it is very possible that additional variables are related to those we are studying, and as yet not made explicit in our theory. Therefore, it is possible to have control over the variables on which we are focusing, the related variables in our theory, and not to have control over still some additional important variables. The best possible treatment would be one for which certain procedures are used to control this last category.

When the research report fails to state a theory, three levels of quality can be described. The lowest is the instance in which something occurred, the details of which are not known. This is the situation experienced by most historical and descriptive researchers and, in such studies, is not bad. In both of these, something has happened to a group which makes them of interest to the researcher. However, that something, those experiences which are the treatment being studied, cannot be delineated by the researcher. All that is known is that something of an undescribed nature was experienced by the subjects studied.

The next level would be the use of a procedure which is generally known in the field but which is not described in detail in the research. An example of this can be found in the numerous studies which refer to "the traditional method" of teaching. Unless that "method" is more carefully defined, the nature of the treatment cannot be considered to have been described in detail.

The third level of quality, given the lack of a theory statement, requires that the treatment employed be described in sufficient detail to enable another researcher to replicate (repeat) the study.

The six items above are the points which make-up the TREATMENT SCALE on the Data Quality Cube.

They are:

T_1 = No theory; something undefined happened to the units studied.

T_2 = No theory; treatment not thoroughly described, used research procedures described elsewhere.

T_3 = No theory; treatment described in detail in the report.

T_4 = Theory stated but no controls on variables.

T_5= Theory stated and mediating variables controlled.

T_6= Theory stated, mediating variables controlled, and techniques used to distribute possible extraneous variances.,

This scale serves as the basis for the labels in the assessment of the treatment in the next section of the profiling flow chart.

Fold out the chart on page 43.

The considerations and decisions necessary to select one of the treatment labels starts at connector number 3. The Profiler is asked first to identify the details of the treatment as specified in the report. Next, he asks, "Has each step of the treatment been specified?" If the answer is negative, still another question is raised about the detailing of major features of the treatment. "Were the major details of the treatment stated or was a standardized procedure used?" Three possible responses exist. If neither of these have been identified in the report the document is labeled on the profiling sheet as T_1 - something of an undefined nature happened to the units studied. If either standardized research procedures were used or the major steps were detailed T_2 is checked on the profiling sheet. In this case the main features of the treatment are known but details necessary for replication may very well have been omitted.

If the answer to the question about specificity of the treatment is affirmative, the Profiler is asked to find three things: (1) the theoretical bases for the treatment; (2) controls for the variables known to be involved in the theory; and (3) controls for those variables extraneous to the theory. Profiling continues through the examination of the question, "Does a theory identify relevant variables and detail their interrelationships?" If the answer to that question is "No," the report is labeled T_3 - no theory is stated but the treatment or experiences of the units studied is described in detail. If the answer to the question was "Yes," still another question is raised. That question seeks information regarding the level of control described in the research. If a theory was stated but controls were asserted over only the variable or variables studied the report is labeled T_4. A T_5 label is attached to those projects which describe controls for the variables being studied *and* those related variables stated in the theory. Finally, the T_6 label is applied to those projects which include controls listed in T_5 plus procedures for controlling variables extraneous to the theory. An example of a procedure used here is the random assignment of units to the various aspects of treatment in the study. If individual units display differences in the characteristics on variables not known to be relevant in the theory, it is assumed that randomization will distribute those differences on a chance basis among the several aspects of the treatment. Once the Profiler has moved to a label for the treatment he is directed by the flow chart to connector number 4. *Remember only one label should be made for each treatment.* That connector leads into the analysis of the quality of the measurement activity in the report.

To measure, according to Webster, is to use a standard, to asertain the extent, degree, quantity, dimensions, or capacity of something. All of these terms connote the use of numerical quantities rather than verbal descriptions. Standard research texts often define MEASUREMENT as the assignment of numbers to objects according to specified rules. That definition suffices well in this context. All research activities, be they historical, descriptive, or experimental, are incomplete without measurement. The historical researcher collects bits of recorded information, classifies them according to rules, and counts their frequency in the established categories. The descriptive researcher collects current bits of information, classifies them. and therefrom generates distributions. The historical researcher classifies *after* finding the bit of information; the descriptive researcher classifies *as* he collects the information; and the experimenter classifies *before* he gathers his data. In all cases, numbers are generated according to pre-specified rules.

It is possible to specify rules, follow them carefully, and still wind up with a poor quality of measurement in a given project. An absurd example helps explain the point. Consider the task of determining which schools in a system have the highest academic achievement level. The data-gatherer is told to stand at the front entrance of a school and count the number of students who have to duck their heads to enter the doorway. That counting task is a very well defined activity of assigning numbers. It does not, however, provide a good measure of the academic achievement of the school. It would not be considered as a valid instrument for measuring achievement. In research this term INSTRUMENT, refers to the tools used in measuring. In a given research project the term "instrument" may refer to a standardized test, a test devised for the project, a questionnaire or some apparatus designed to make a record of an event or performance. To validly measure achievement, the researcher would have to scrap his counting of students who must duck to get through the door, and utilize an instrument which gets at achievement. VALIDITY, then, is the degree to which an instrument is a true measure of those items being investigated.

Another characteristic of sound measurement is RELIABILITY. If the use of a measuring instrument produces one score at one point in time and another score at another point in time, it cannot be considered as a good instrument. Consider again an absurd example. A rubber yard stick would not serve as a reliable measure of length when used by different persons or by the same person at different times to measure the length of a room. Such an instrument would yield different scores. A reliable instrument is one which produces identical scores when used by different persons or the same person at different times to measure the same item.

Still one further item is needed to consider the quality of measurement. That item is OBJECTIVITY. It also concerns the degree to which different people would obtain the same result. It focuses on differences caused by the *user* where reliability focused on differences caused by the structure of the tool itself. Consider in this case the task of measuring the quality of a musical performance. Although one might be given a very specific set of rules, there is still a great degree of personal judgement involved. That fact makes it difficult for a number of judges to record the same scores for the same performance. An example of a measuring instrument with high objectivity would be a commercially standardized test for which the correct response to each item is given and explicit details are provided about computing a score. In such a case, every one who grades the test ought to be able to arrive at the same score for any given answer sheet.

In summary, then, total evaluation of the data in a given piece of research must consider the validity, reliability, and objectivity of the measurement activity. If the data upon which the conclusion is to be based are generated by valid, reliable, and objective measuring activities, the measurement aspect of the research project can be considered to be methodologically sound.

In light of the preceding discussion, six levels of measurement data quality can now be described. The two extremes, as indicated in an earlier section, were: (1) a low, representing the obtaining of data with instruments for which we have no information (The data were probably from records kept for other purposes and merely "reworked" for this study.); and (2) a high, representing measurements obtained through perfectly valid, reliable, and objective techniques (Probably using, but not restricted to, a well-known commercially produced and standardized instrument). Between these points lie four additional quality levels that are predicated on two concepts - how the instrument was developed and its relative strength for the specific measurement task undertaken.

Instruments are born out of one of three sources - a new project, an earlier project, or a test manufacturing company. For this presentation these have been designated as PROJECT DEVELOPED, OTHER PROJECT DEVELOPED, and COMMERCIALLY PRODUCED instruments respectively.

The six levels on the MEASUREMENT SCALE are:

M_1= Information available that the instrument is INVALID for this use.

M_2= Project Developed instrument with LOW validity (V), reliability (R), objectivity (O), or NO INFORMATION about Commercially Produced and Other-Project Developed instrument.

M_3= Used Commercially Produced or Other-Project Developed instrument with LOW V, R, O for this application.

M_4= Used Project Developed instrument or Other-Project Developed instrument with MODERATE V, R, O for this application.

M_5= Used Commercially Produced instrument with MODERATE V, R, O or other instrument with HIGH V, R, O for this application.

M_6= Used Commercially Produced instrument with HIGH V, R, O for this application.,

The rank-order of the measurement scale reflects an inherent assumption. It is that the greater the professional exposure of an instrument, the greater the probability of its criticism, revision, and validation. Hence, commercially produced, commercially standardized instruments *should* be the best available. Next come instruments that have been tested on other projects. Last in line are instruments whose first wide-spread exposure occurred with the publishing of the report being evaluated.

Connector number 4 in the upper left hand corner of the flow-chart on the measurement fold out (page 44) leads into a consideration of the quality of measurement. In recognition that a number of measuring instruments may be employed in any given study, the Profiler is asked first to identify each one and then to list it on the back of the Research Profile Sheet. *The profiling task proceeds on a cyclical basis until all of those listed instruments are labeled.*

The first step in the cyclical labeling of the individual instruments is a question which separates data-gathering instruments, and associated data about their validity and reliability from those with high PFMA (Plucked From Mid-Air) factors. Those instruments that were just "plucked" and forgotten have a termination label of M_2, a low level on the measurement scale. A "Yes, But Invalid" answer, however, is worse than having no information at all. Label it M_1. Remember -- The instrument must be a true measure of at least those items being investigated. If no label has been reached yet, continue through the chart at the "Yes, Valid" exit.

Determining the origin of the instrument is the next concern. Responses to the question, "Is the instrument Project Developed (PD), Other-Project Developed (OPD), or Commercially Produced (CP)?" will be more fully explained in the corresponding paragraphs below. The Profiler selects the answer appropriate for the instrument being evaluated and applies the instructions under that answer. When the point of labeling is reached, recycle and follow the *same* procedure for the next instrument listed on the back of the profile sheet.

Project Developed (PD). If the measuring instrument was developed expressly for this project, the answers to the question, "Is the Validity, Reliability, Objectivity (V,R,O) low, moderate, or high?" carry a slightly lower weighting than for other instruments. This same question is asked for the other two categories also. "Low" has an M_2 level, while a "Moderate" V,R,O for tests from this project are placed at the M_4 level. A project developed instrument with "High" ratings is given M_5. Again, the rationale for giving a less-than-the-best rating here is that the top is reserved for commercial products with their greater audience of critics.

Other-Project Developed (OPD). The validity, reliability, and objectivity (V,R,O) of instruments adopted from other research projects must now be examined. Lack of a valid connection between the instrument and its use in this study warrants an M_1 level. If there is no information other than its origin, and M_2 is listed.

Since an OPD instrument, by definition, has undergone previous development and testing, a "Low" V,R,O has been rated M_3, the same as for a commercial product. For the "Moderate" and "High" responses the instrument is treated the same as a new one created just for this project; and M_4 and M_5 are assigned, respectively.

Commercially Produced (CP). Still another tack is taken for test which are commercially developed, produced, and standardized. While lack of information and lack of substantiation of the instrument's applicability rate an M_2 or M_1, it is the rating of the validity, reliability, and objectivity levels which reflect the greater confidence in the test manufacturers, like ETS, California Test Bureau, etc. The "Low," "Moderate," and "High" V,R,O designations yield M_3, M_5, and M_6.

The Profiler will note that a commercially standardized test which fails to give validity, reliability, and objectivity information for this application is considered of greater quality than is a project developed instrument with the same lack of information. This ranking is based on the assumption that measurement specialists were involved in the development of commercially standardized instruments. And, that since they are offered on a continuing basis by organizations specializing in measurement, they are more likely to be sound than are project developed ones. In the latter case, individuals frequently are involved who do not have as thorough an understanding of the subtleties of measurement as do the people employed by commercial producers.

At this point it should be restated that this labeling activity is repeated for each measuring instrument employed in the research project. When the final instrument is labeled, the individual is directed to connector number 5, the start of the statistical analysis evaluation.

The Data Quality section of the Research Profiling Flow Chart has now been completed. It should be refolded before going on to the Statistical Analysis of Data which follows.

THE STATISTICAL ANALYSIS OF DATA

STATISTICAL ANALYSIS, in the context of these profiling *instructional materials, is the process of simplifying* collected data. In most research efforts a large quantity of data is generated. In their raw form data often defy interpretation. A statistical analysis is usually performed to facilitate interpretation. For example, consider ten students who have taken the same test. The number of correct responses for each is: 24, 18, 17, 26, 14, 28, 25, 26, 21, and 20. As they are presented here, it is difficult to see them as anything other than an assortment of individual numbers. Most people would not bother to even finish reading all ten numbers. But, analysis can make them understandable as a group. Analyses range in complexity from the simple ordering of the numbers (14, 17, 18, 20, 21, 24, 25, 26, 26, 28) to the determination of central tendency (21.9) and dispersion (16.4 - 26.4). With still more information about the data it would be possible to show measures of association and even make inferences and predictions.

The first of these, ORDERING, shows us that the scores range from a low of 14 to a high of 28, and that two of the scores are identical at 26. The measure of CENTRAL TENDENCY called the mean, or average, equals 21.9. Finally, DISPERSION is the manner in which these numbers spread on either side of the central point. These statements have more descriptive value than does the jumble of numbers presented earlier.

Along with simplifying sets of numbers, statistical analyses are undertaken: (1) to describe a group on one or more variables; (2) to determine whether different kinds of data increase and decrease together; and (3) to determine the amount of confidence that can be placed on the generalizability of observed data. The first of these is generally called DESCRIPTIVE STATISTICS. The second is CORRELATIONAL or ASSOCIATIONAL STATISTICS. And the third is INFERENTIAL STATISTICS. The number of different statistical analyses that can be performed under each of these categories is quite large. The selection of a specific statistical technique is dependent upon: (A) the general purpose of the analysis (the three categories listed immediately above; (B) the scalor nature of the numbers involved; and (C) the number of variables in a specific study. (A) and (B) are discussed in detail below before returning to the Research Profiling Flow Chart. The number-of-variables criterion (C) should be self-explanatory.

The SCALOR NATURE criterion (B) is concerned with both categories of variables and levels of scales. Three categories of variables can be described:

(1) CONTINUOUS VARIABLES (For example: number of correct responses on a test, age, number of years of schooling, standardized test scores.)

(2) DICHOTOMOUS VARIABLES (Items that are either a or b, such as. sex - male or female, in or out of school, answered or failed to answer a test question.)

(3) ARTIFICIALLY DICHOTOMOUS VARIABLES (For example: number of persons who are over and under age 21, number who passed or failed an entire test.)

Connecting categories of variables and levels of scales are the four cumulative properties of numbers that follow:

(1) NAME CONSTANCY - Each number serves as the name of a distinct group. (Three is a name on a distance scale which refers to a precise category of distances, five refers to still another category of distance.)

(2) ORDER - Different numbers fit together in a recognized sequence. (Two comes after one and before five. Five comes somewhere after two and before sixteen. A distance scale marked one, two, sixteen, five would be incorrect. This is in contrast with the use of numbers as names in situations where order is not important, as with the numbers on football players' jerseys.)

(3) EQUALITY - Differences between adjacent numbers on a scale are equal. (The distance on a ruler between the numbers 2 and 3 is identical to the distance between the number 8 and 9. In contrast, attitude measurement typically is based on the following scale: (1) Strongly agree, (2) Agree, (3) Neutral, (4) Disagree, (5) Strongly disagree. *The change in attitude from (1) Strongly* agree to (2) Agree is not necessarily the same as the change from (2) Agree to (3) Neutral.)

(4) ZERO POINT - A true zero point must exist. (On a distance scale there is such a thing as *no distance* a zero point. On some scales that point does not exist, there is no such thing in the *living* human as zero intelligence.),

These four properties are cumulative. That is, property number two, order, cannot exist if property number one, name constancy is not present. Similarly, property number four requires the existence of one, two, and three.

Given these properties and their cumulativeness, four general levels of scales can be described. They are:

(1) NOMINAL SCALES - the use of numbers as names (Example: assigning numbers to students of different nationalities for analytical purposes; Canadian - 1, English - 2, French - 3, Mexican - 4, etc.)
(2) ORDINAL SCALES - the use of numbers as names for categories that have an inherent order. (Example: rank in class-first, second, third...)
(3) INTERVAL SCALES - the use of numbers to indicate equal spacing between ordered *and* named categories. (Examples: intelligence tests, time.)
(4) RATIO SCALES - the use of numbers as names of equally distant units on measures that have a zero point. (Examples: number of correct responses on a specific test.),

Discrimination among these scalor levels is necessary for proper selection of statistical procedures. It does not make sense, nor does it follow empirically, that a person who ranks 5th in his calss on math grades, 3rd on English grades, 8th in science, and 4th in social studies will rank 20th (their sum) or 5th (their average) on a composite of those ranks. It is recognized that such *inappropriate* mathematical manipulations can be made; however, it should also be recognized that in so doing, the nature of the statistical conclusion has been changed. A rank number is information about one's performance in relation to others. An average of several ranks has no relationship to performance. It is merely an exercise in the addition and division of numbers. Statistical formulations have been devised which are appropriate for data generated through the use of the scales above. The researcher's task, then, is two-fold. He must determine the characteristics of his data *and* find the statistic formulated on the same principles.

The term DESCRIPTIVE STATISTICS includes analytic procedures developed as aids in describing a population or a sample of a population. The term generally covers measures of central tendency and measures of dispersion. Measures of CENTRAL TENDENCY are the statistics called the mean, median, and mode. The MEAN is the arithmetic average of a group of numbers or quantities. It is found by summing the numbers and dividing by how many numbers there are. When the numbers refer to ranks of individuals or order of items, typically a median is calculated. A MEDIAN is a point above and below which fifty per cent of the individuals or items fall. MODE is used for categorical information and refers to that category which has the largest number (Highest frequency) of entries.

DISPERSION includes standard deviation, semi-interquartile range, and total range. A STANDARD DEVIATION is a restricted type of range. It is the distance above or below the mean that is necessary to encompass 34 per cent of a normally distributed population. The SEMI-INTERQUARTILE RANGE is the distance above and below the median which is required to encompass 50 per cent of the units. RANGE is the distance between the upper- and lower-most scores (the extremes) in a group.

CORRELATIONAL or ASSOCIATIONAL STATISTICS are appropriate to those investigations where it is desirable to indicate whether high scores on one measure would coincide with high scores on another. For example, what is the likelihood that students who have high measured intelligence will also have high scores on a specific achievement test and vice versa? This analytic purpose is met through the computation of either correlation or association by a number of statistical procedures (Including: Pearson r, Biserial r, point-biserial r, Phi coefficient, Kendall's W, coefficient of concordance C, Spearman's rho, multiple r, etc.). In the selection of a correlational statistic the individual considers first the *number* of variables. In some cases correlation involves two variables; in others three or more. Next, he obtains information about the scalor nature of the variables involved thereby leading to a specific statistic to be used. More explicit directions for selecting the correct associational statistic will be presented in the flow chart section dealing with statistical analysis.

A basic requirement for computation of associational statistics is the existence of a common tie among the scores on the variables to be correlated. For example, to calculate correlation between intelligence and achievement, scores on *both* variables must be available on the *same* individuals. It *cannot* be calcualted if one group's intelligence scores are known and another group's achievement scores are known.

A person makes an INFERENCE when he uses something he has observed as evidence about something else. Many situations in education call for inferences. An educator may want to know how intelligent a student is. It is impossible to look directly at the student's intelligence by opening his head and examining the matter there. Since that cannot be done, something else is observed that *can* be logically accepted as an indicator of intelligence. Typically, the student is asked to do something. If his response fits a predicted pattern, it is inferred that he is intelligent. In a research project that attempts a "test of a hypothesis" the investigator is in essence saying, "If this hypothesis is a ture statement, specific things should happen that can be observed." He constructs a situation in which those things ought to occur, measures to see if they did occur, and uses those measurements (numbers) as the basis for inferring that his hypothesis is (or is not) true. INFERENTIAL STATISTICS are procedures that have been developed to simplify numbers used as the basis of inferences.

As in the cases of descriptive and associational statistics, a large number of inferential statistics exist. Again, choice of a specific statistic is dependent upon the scalor nature of the measuring instruments, the number of variables involved, the number of groups involved, and the specific aspect of the data being analyzed. Further explanation of these determinants can be found in any one of numerous statistics texts (Guilford, 1965; Senders, 1958; Walker & Lev, 1953; etc.).

Thus far, this discussion has covered the following aspects of the research process: the logical argument (hypothesis testing and answering empirical questions) and the data quality (unit representativeness, treatment details, and measurement instruments).

Given and understanding of those concepts, we are ready to return to the Research Profiling Flow Chart. The final evaluation activity covers the data analysis procedures used in the research study being evaluated. It begins at connector number 5 on the ANALYSIS fold out (page 45).

As was the case in the area of measurement, it is not uncommon for a specific project to employ several statistical techniques in the data analysis. The Profiler identifies each statistical technique and lists it on the back of the profile sheet. For each procedure listed, he follows the directions of the flow chart until a label is attached.

The pathway to a label starts with a question about the purpose of the analysis. At that point three branches are shown. One of these branches is followed until a lable is identified. The directions for this labeling have been structured in a way which requires that the Profiler read only the material related to the branch taken. The three branches are keyed to answers to the analysis purpose question and are: (A) To describe; (B) To show association; and (C) To support an inference. The necessary material for each of these branches is stated below under the three headings.

(A) TO DESCRIBE. If the purpose of the analysis was to describe, the Profiler asks a second question, "Are data based on a nominal, ordinal or interval scale?" The answer to that question names a row in Chart A. Each cell in that row contains a statistic that *should* be in the report. Those that are not are to be listed on a profile sheet.

(B) TO SHOW ASSOCIATION. If the purpose of the analysis is to show association, that is the degree to which scores correlate, a distinction has to be made as to whether two or more variables are involved. If the answer is "TWO," the Profiler determines the categories of the two variables and checks Chart B for the appropriate statistic. The labels of the COLUMNS in Chart B are the categories that underly Variable 1. Given a specific variable one of these columns is identified. The same holds for Variable 2 and the ROWS of Chart B. Using the respective categories for the *two* variables, the cell at the intersection of that row and column contains the statistic that should have been utilized in the report. If it was, the label "A" for Appropriately Analyzed is checked on the profile sheet. If the category of the data is too low to warrant using the statistic reported in the study, label it "I" for Inappropriately Analyzed. Example: Analyzing dichotomous data with a Pearson r when the data should be continuous. If the data fit these categories but none of these statistics is shown in the report, indicate an "M" for Missing.

If there are *more than two* variables, appropriate two-variable correlations ought to be found in the report (follow the procedures described immediately above). In addition, check the appropriate special cases listed under Chart B. Again, if these analyses are there, the report is labeled "Appropriately Analyzed"; if not, "Inappropriately Analyzed."

(C) TO SUPPORT AN INFERENCE. If the statistical analysis is undertaken for the purpose of inference, the Profiler is asked to label the independent and dependent variables in the study.

The INDEPENDENT VARIABLE is that variable which is manipulated by the researcher. The DEPENDENT VARIABLE is the variable which is expected to change *as a result* of that manipulation. Consider, for example, the hypothesis that teaching method A will produce greater achievement than teaching method B. The independent variable, the variable to be manipulated in this case, is instructional method. It has two categories, Teaching Method A and Teaching Method B. The dependent variable is achievement. That is, the level of achievement is said to be dependent upon the aspect of instructional method used in the study.

Once the variables are categorized as dependent or independent, the Profiler turns to Chart C to determine the appropriate statistic. He does this by identifying the number of dependent variables (1 or > 1), the number of independent variables (1 or > 1), and their scalor nature. Using that information he can find the appropriate row and column on

Chart C. The statistic stated within that row and column is appropriate for the analysis. If it matches with the statistic actually used by the researcher, the report is labeled "Appropriately Analyzed"; if not, it is labeled "Inappropriately Analyzed." When all the statistical procedures have been labeled the evaluation of the project has been completed and a summary profile on that study has been created.

BIBLIOGRAPHY

Campbell, D. and Stanley, J., "Experimental and Quasi-Experimental Designs for Research on Teaching." In HANDBOOK OF RESEARCH ON TEACHING. N. L. Gage Ed. Chicago: Rand McNally. 1963 p. 175-190.

Gephart, W. J., and Ingle, R. B., EDUCATIONAL RESEARCH: SELECTED READINGS. Columbus, Ohio: Charles E. Merrill Publishing Co., 1969.

Guilford, J. P., FUNDAMENTAL STATISTICS IN PSYCHOLOGY AND EDUCATION: FOURTH EDITION. New York: McGraw Hill Book Company, 1965.

Kerlinger, F. N., FOUNDATIONS OF BEHAVIORAL RESEARCH. New York: Holt, Rinehart and Winston, Inc. 1964.

McNeil, J. O., "Programed Instruction Versus Usual Classroom Procedures in Teaching Boys to Read." AMERICAN EDUCATIONAL RESEARCH JOURNAL 1:113-19; March 1964.

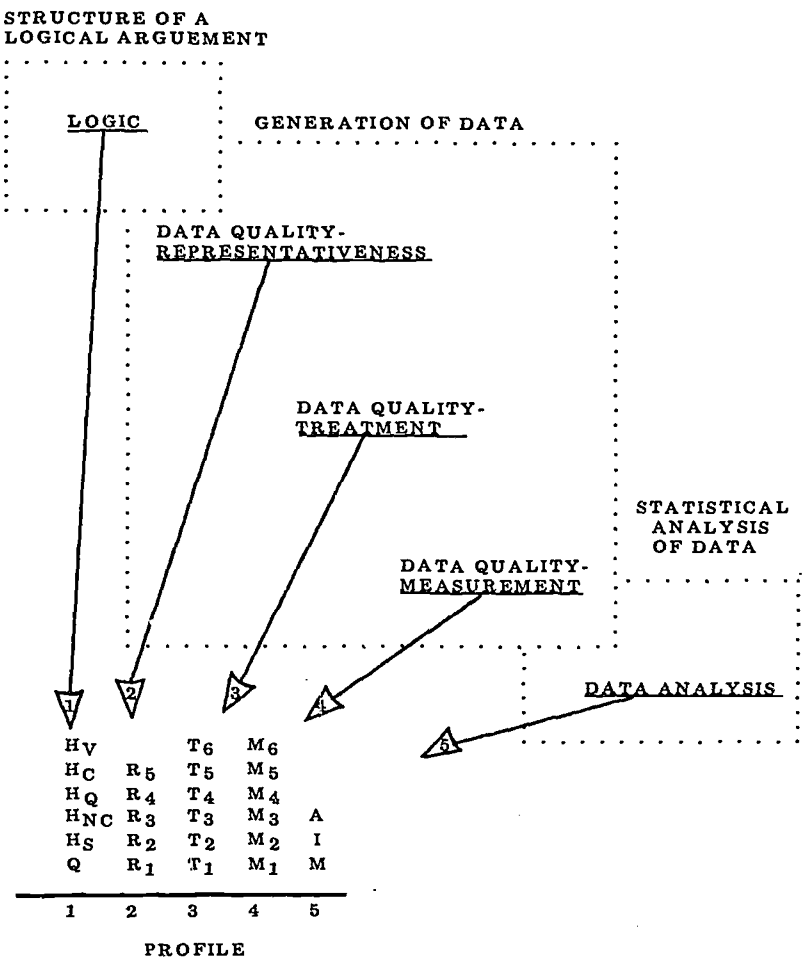
Polya, G., PATTERNS OF PLAUSIBLE INFERENCE: MATHEMATICS AND PLAUSIBLE REASONING Volume II, Princeton, New Jersey: Princeton University Press, 1954.

Raths, J., "Plausible Logic in Educational Research" in EDUCATIONAL RESEARCH: SELECTED READINGS. (Eds.) W. J. Gephart and R. B. Ingle, Columbus, Ohio: Charles E. Merrill Publishing Co. 1969.

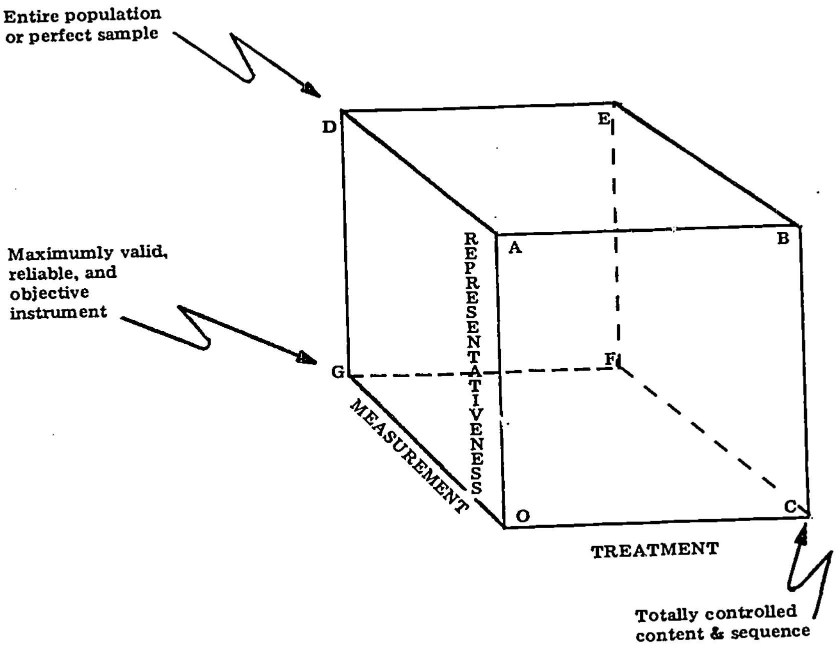
Senders, V., MEASUREMENT AND STATISTICS. New York: Oxford University Press, 1958.

Walker, H. M., and Lev, J., STATISTICAL INFERENCE. New York: Henry Holt and Company, 1953.

FACETS OF THE RESEARCH PROCESS

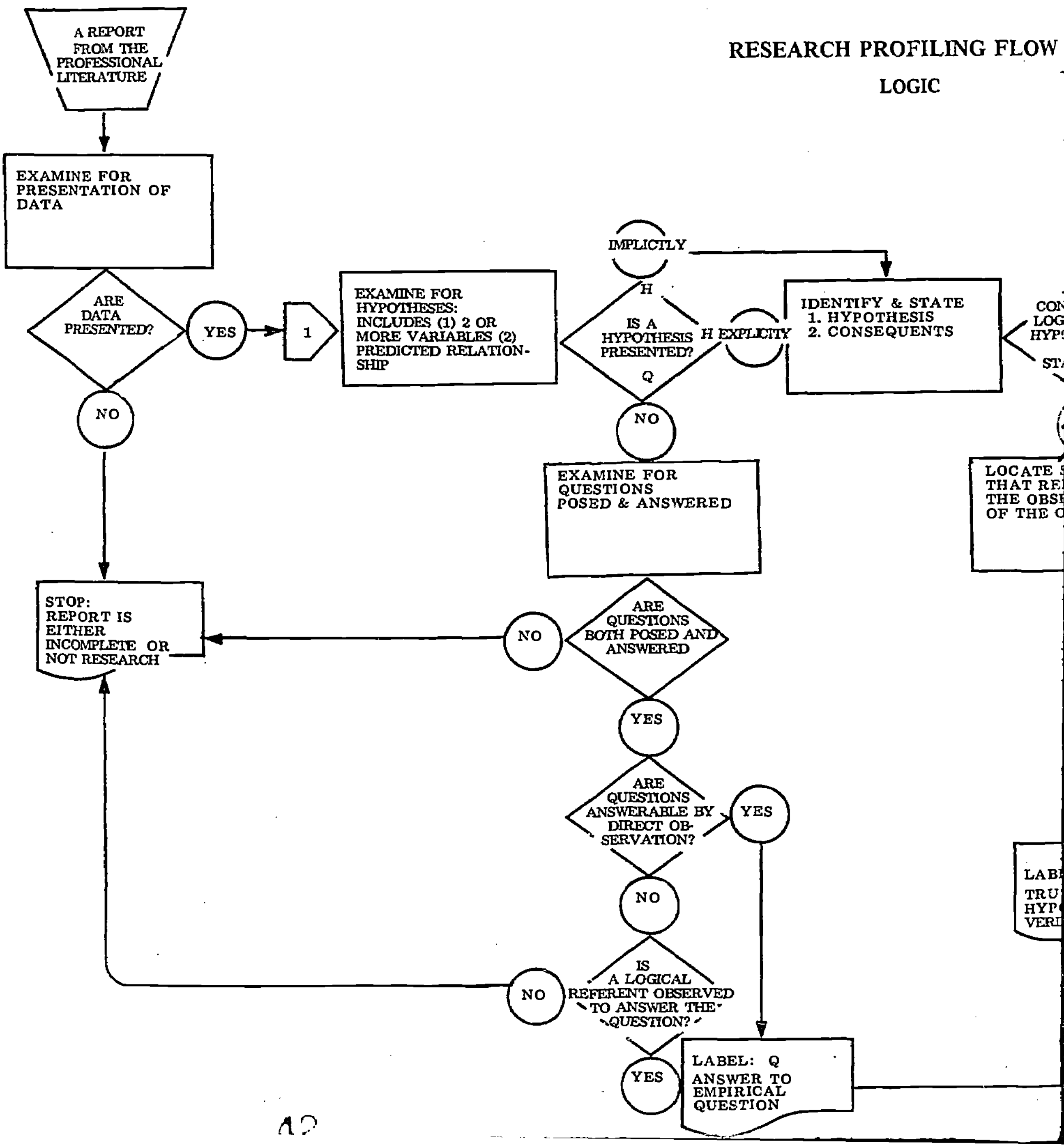


THE DATA QUALITY CUBE



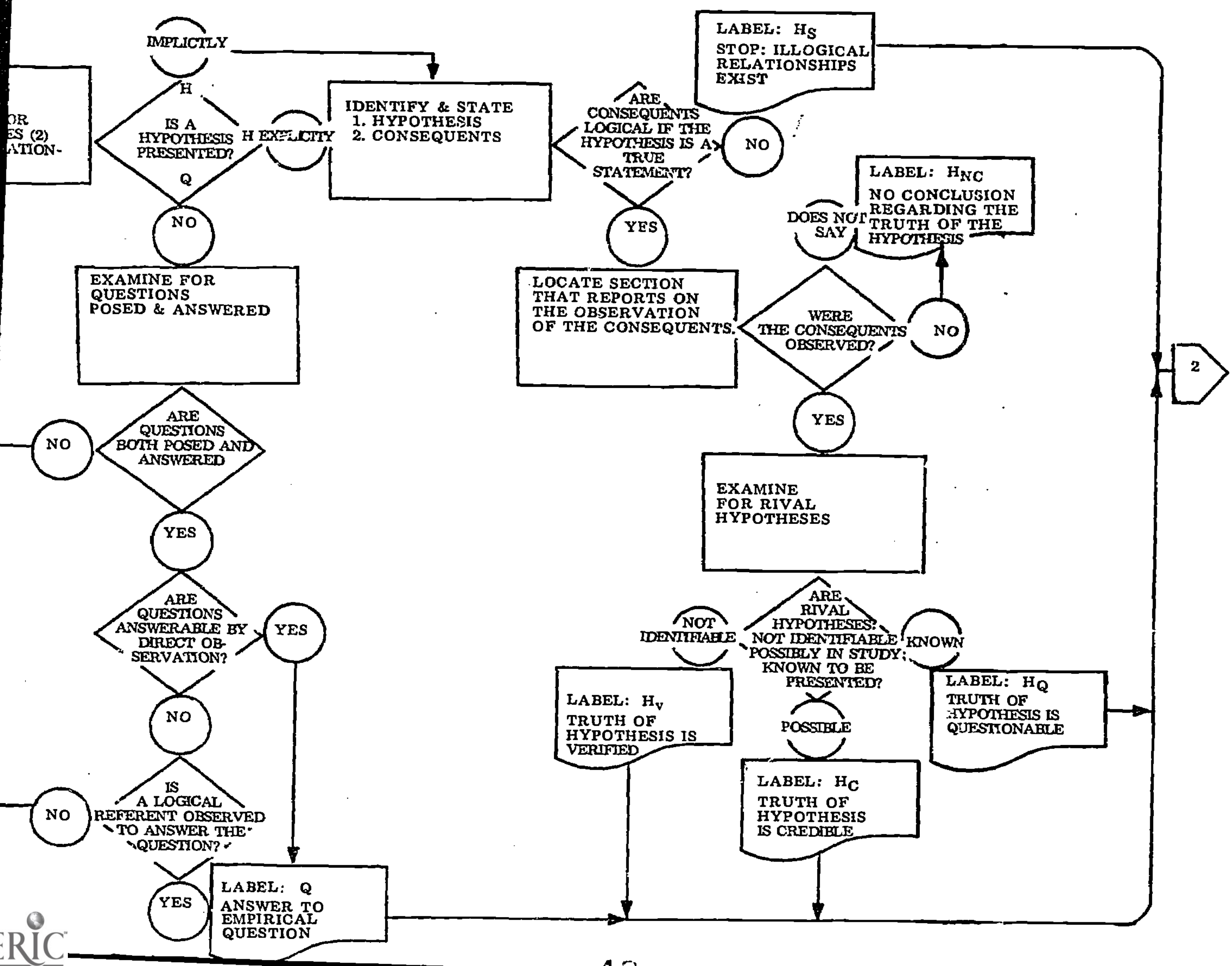
RESEARCH PROFILING FLOW

LOGIC



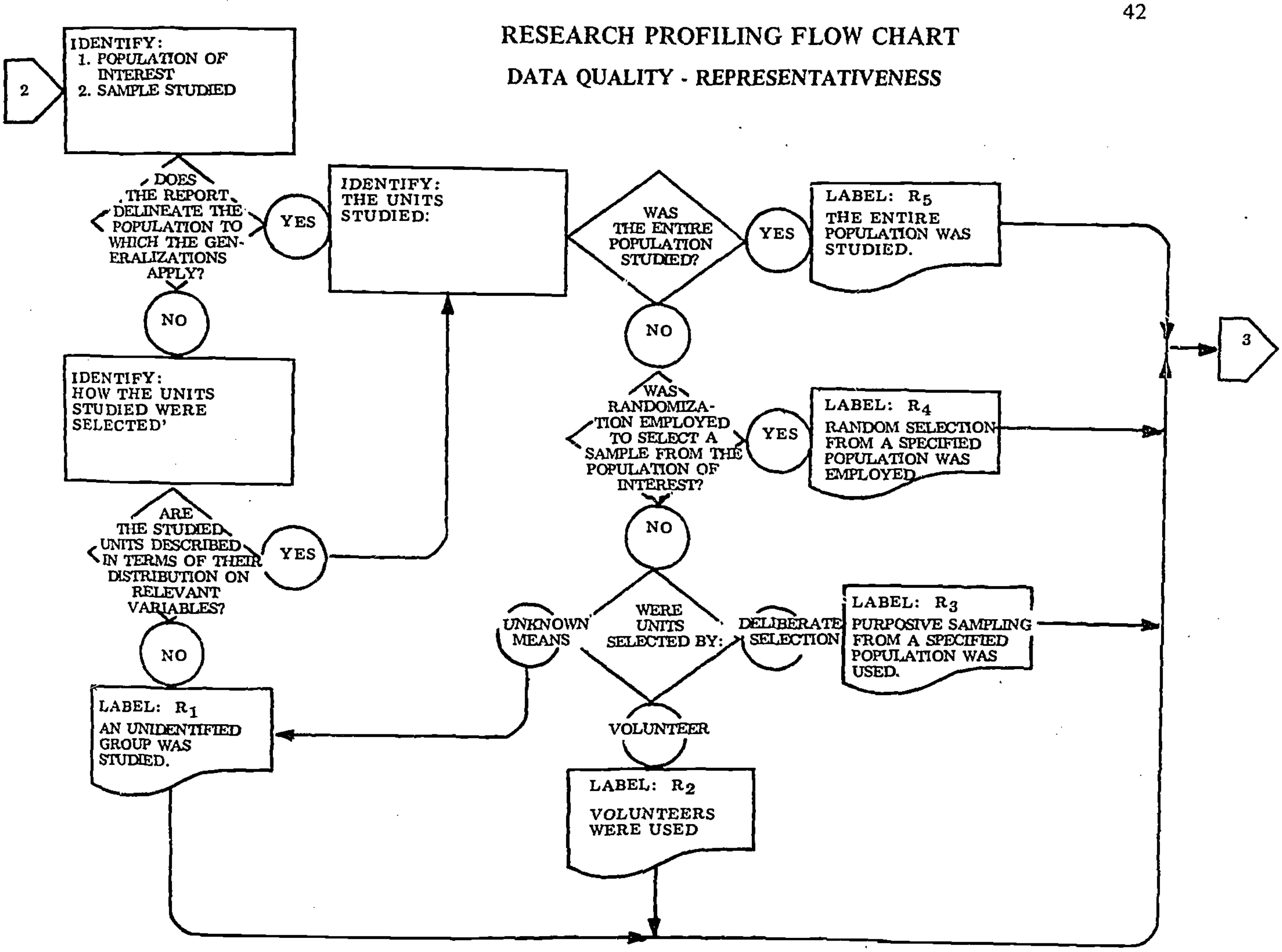
RESEARCH PROFILING FLOW CHART

LOGIC



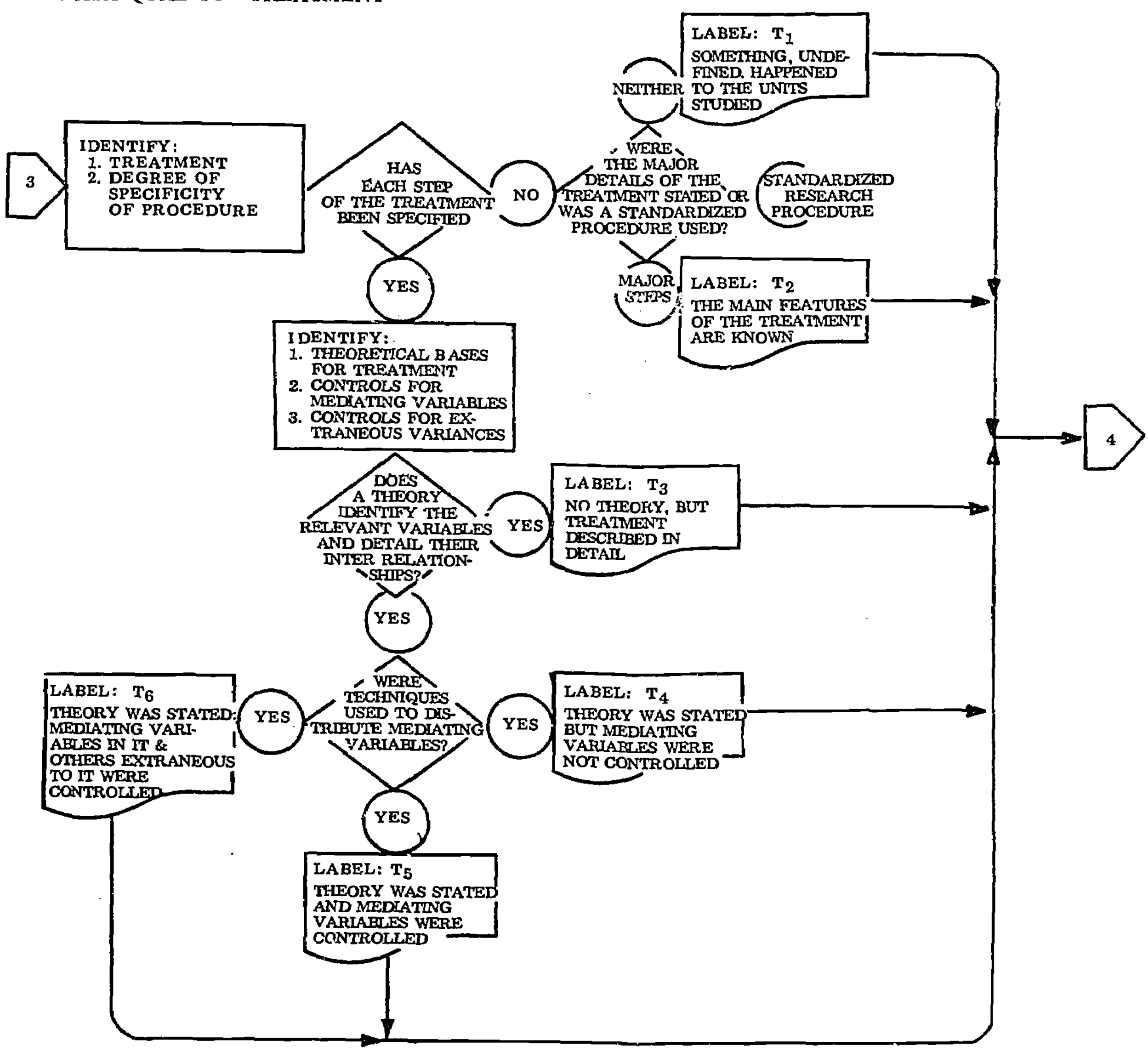
RESEARCH PROFILING FLOW CHART

DATA QUALITY - REPRESENTATIVENESS



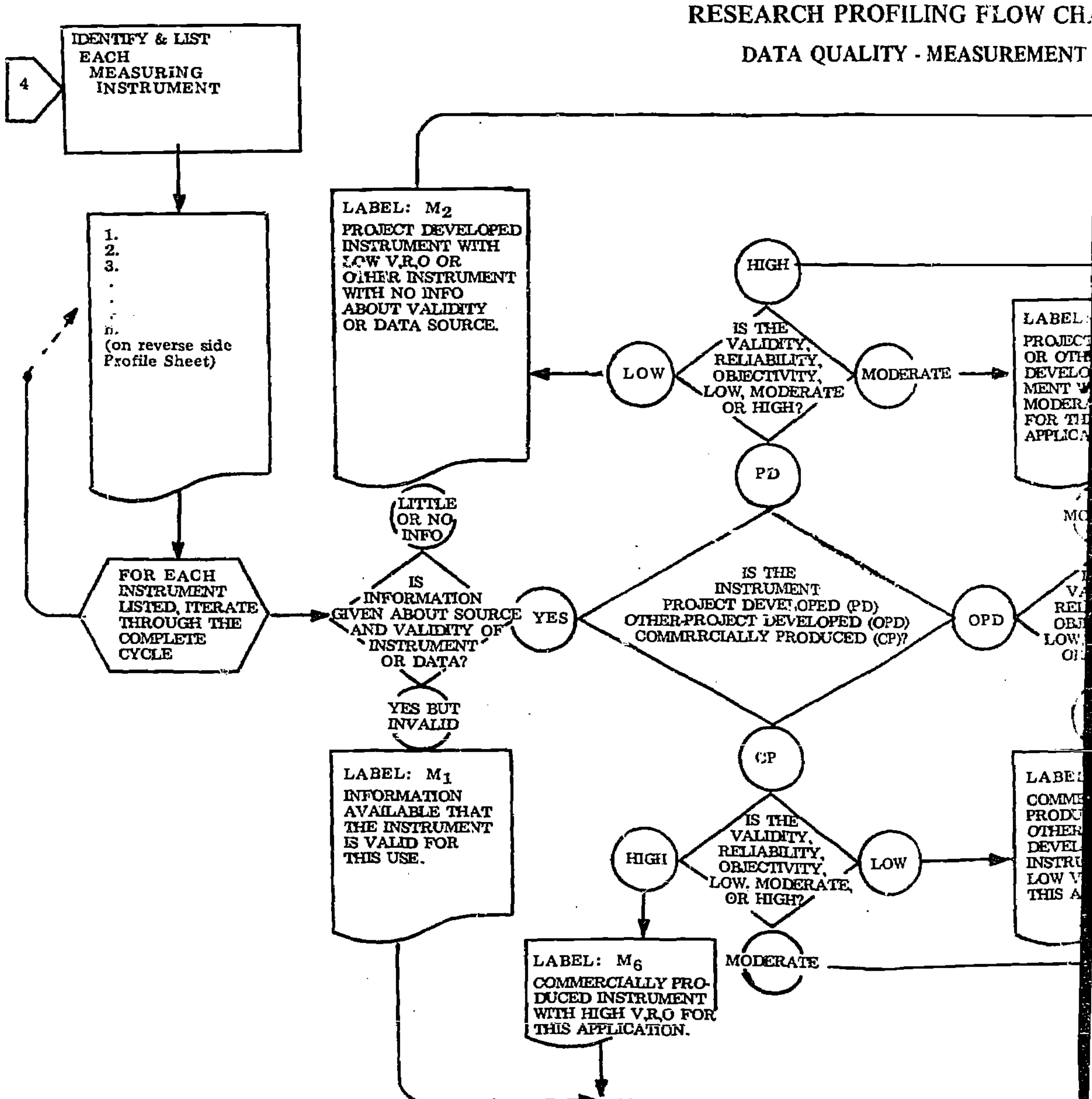
RESEARCH PROFILING FLOW CHART

DATA QUALITY - TREATMENT



RESEARCH PROFILING FLOW CH

DATA QUALITY - MEASUREMENT



RESEARCH PROFILING FLOW CHART

DATA QUALITY - MEASUREMENT

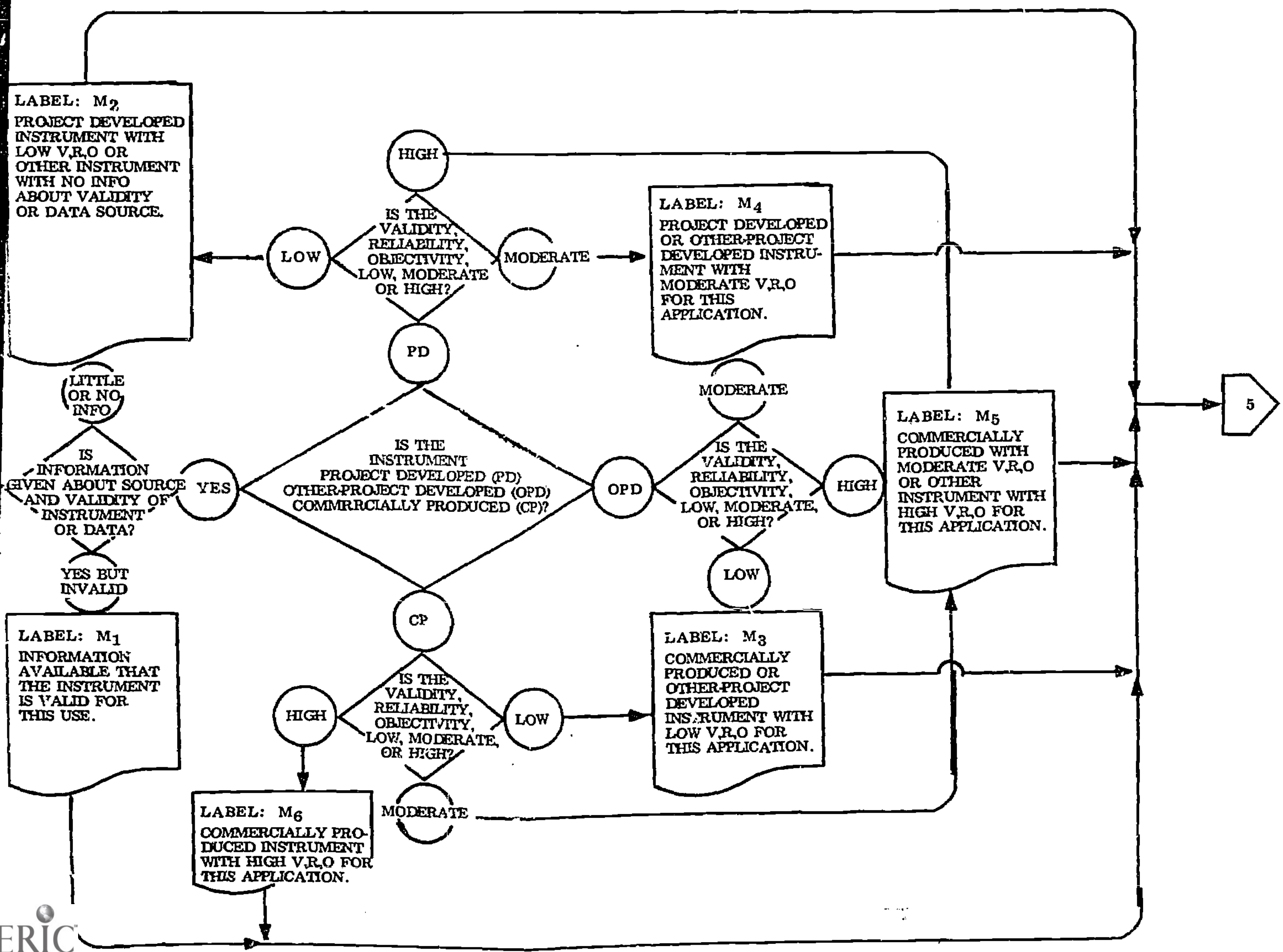


CHART A

POPULATION DESCRIPTORS

	DISTRIBUTION	CENTRAL TENDENCY	DISPERSION
NOMINAL	FREQUENCY IN EACH CATEGORY	MODE	
ORDINAL	FREQUENCY IN EACH SCALOR POSITION	MEDIAN	SEMI-INTERQUARTILE RANGE
INTERVAL/ RATIO	FREQUENCY IN EACH INTERVAL	MEAN	STANDARD DEVIATION

CHART B - MEASURES OF ASSOCIATION

VARIABLE 1*

	CONTINUOUS	FORCED DICHOTOMY	DICHOTOMY
CONTINUOUS	PEARSON r	BISERIAL r	POINT BISERIAL r
FORCED DICHOTOMY	BISERIAL r	TETRACHORIC r	(NONE AVAILABLE USE CHI SQUARE)
DICHOTOMY	POINT BISERIAL r	NONE AVAILABLE	FOUR FOLD r OR PHI COEFFICIENT

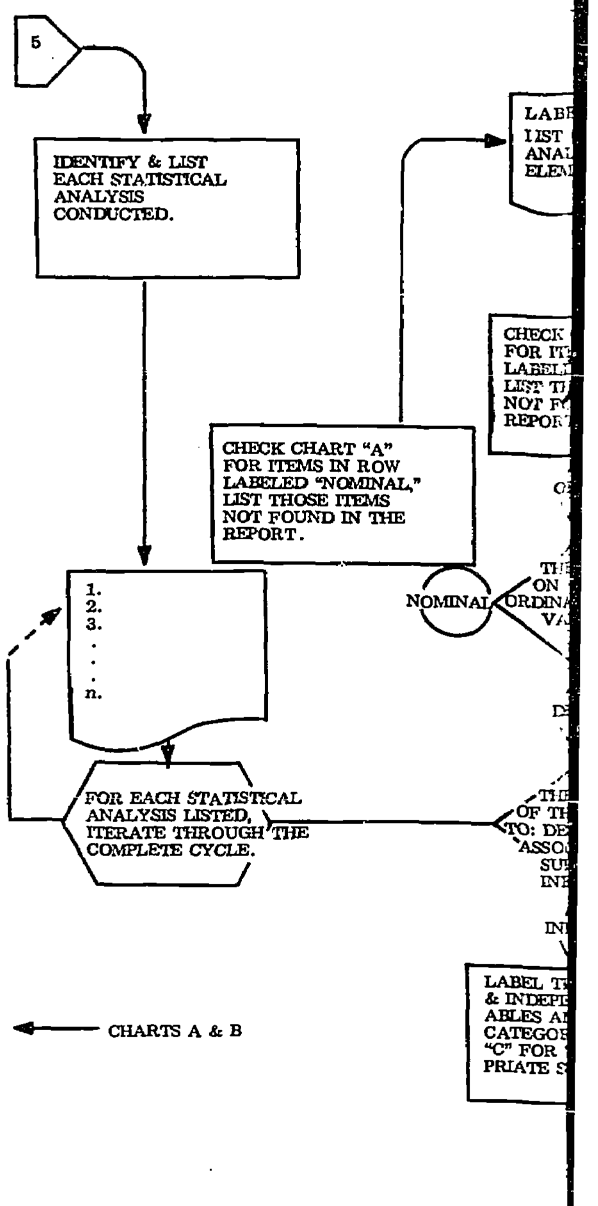
* INTERVAL DATA

SPECIAL CASE
2 VARIABLES

RANK DATA SPEARMAN'S RHO

SPECIAL CASE
MORE THAN 2 VARIABLES

INTERVAL DATA MULTIPLE R
ORDINAL DATA KENDALL'S W
NOMINAL DATA CONTINGENCY COEFF. C



RESEARCH PROFILING FLOW CHART

ANALYSIS

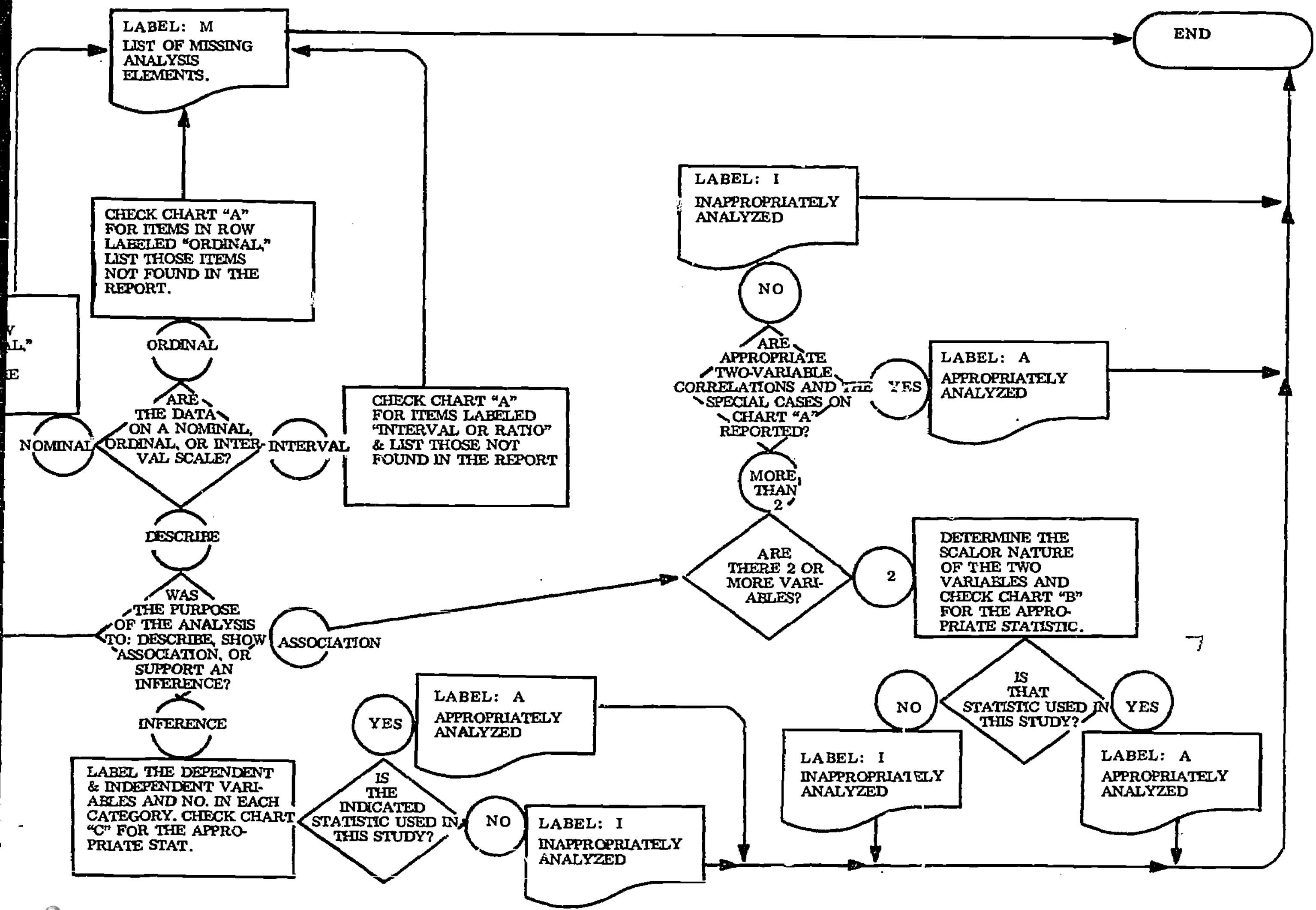


CHART C

INDEPENDENT VARIABLE(S)

DEPENDENT VARIABLE(S)

	NOMINAL 1	NOMINAL>1	ORDINAL 1	ORDINAL>1
NOMINAL 1	FISHER'S EXACT PROB. FOR 2x2 TABLE MC NEMAR'S TEST FOR SIGNIFICANCE OF CHANGES COCHRAN'S Q TEST FOR SEVERAL RELATED PROPORTIONS CHI-SQUARE TEST FOR INDEPENDENCE METHODS FOR MAXIMIZING PROBABILITY OF CORRECT CLASSIFICATION		SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST	
NOMINAL>1			FRIEDMAN'S 2-WAY ANOVA	
ORDINAL 1	SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST KRUSKAL-WALLIS 1-WAY ANOVA KOLMOGOROV-SMIRNOV 1-SAMPLE TEST	FRIEDMAN'S 2-WAY ANOVA	INDEX OF ORDER ASSOCIATION	ANALYSIS OF VARIAN WITH TREND ANALY
ORDINAL>1				
INTERVAL 1	ANALYSIS OF VARIANCE	INDEPENDENT t ANALYSIS OF VARIANCE ANALYSIS OF COVARIANCE	ANALYSIS OF VARIANCE	
INTERVAL>1	CORELATION FACTOR ANALYSIS MULTIPLE DISCRIMINANT ANALYSIS MULTIPLE REGRESSION ANALYSIS	MULTIPLE DISCRIMINANT FUNCTION		

Most of the measures shown are located with reference to the lowest order of data that should be used with them. One should always be able to transform the observed data downward i.e. interval can be considered ordinal or nominal, ordinal can be consi... red nominal.

CHART C

INDEPENDENT VARIABLE(S)

	ORDINAL 1	ORDINAL > 1	INTERVAL 1	INTERVAL > 1
SIFICATION	SIGN TEST MEDIAN TEST MANN-WHITNEY U TEST		ANALYSIS OF VARIANCE	HOTELLING'S T MAHALANOBIS' D^2 FISHER'S DISCRIMINANT FUNCTION
	FRIEDMAN'S 2-WAY ANOVA		ANALYSIS OF VARIANCE CORRELATION FACTOR ANALYSIS MULTIPLE DISCRIMINANT ANALYSIS	RAO'S V_k MULTIPLE DISCRIMINANT FUNCTION
VA	INDEX OF ORDER ASSOCIATION	ANALYSIS OF VARIANCE WITH TREND ANALYSIS		
CE	ANALYSIS OF VARIANCE		REGRESSION ANALYSIS	MULTIPLE REGRESSION ANALYSIS
T FUNCTION			MULTIPLE REGRESSION ANALYSIS	CANONICAL CORRELATION

he
ould
in-

This table is an adaptation of Tatsuoka and Tiedeman's Table 1 in "Statistics As An Aspect of Scientific Method in Research on Teaching" pgs. 154-155, in HANDBOOK OF RESEARCH ON TEACHING, N. L. Gage (Ed); Rand McNally, Chicago, 1963, 1218pp.

H_V		T_6	M_6	
H_C	R_5	T_5	M_5	
H_Q	R_4	T_4	M_4	
H_{NC}	R_3	T_3	M_3	A
H_S	R_2	T_2	M_2	I
Q	R_1	T_1	M_1	M
1	2	3	4	5

RESEARCH PROFILE SHEET

REPORT TITLE: ____________________________

AUTHOR ______________ SOURCE: ______________

____ STOP The report is either not research or it is an incomplete part of the research process.

1 LOGIC

- ____ Q Answer to an Empirical Question
- ____ H_s Stop, Illogical relationship in the test of the hypothesis.
- ____ H_{nc} No conclusion can be reached from this test of the hypothesis.
- ____ H_q Hypothesis is questionable. (Rival hypotheses <u>must</u> be considered a cause of the consequents)
- ____ H_c Hypothesis is credible. (Rival hypotheses <u>may</u> be considered a cause of the consequents)
- ____ H_v Hypothesis is verified. (Rival hypotheses <u>cannot</u> be considered as a cause of the consequents)

2 DATA QUALITY - REPRESENTATIVENESS

- ____ R_1 An unidentified group of subjects was studied.
- ____ R_2 Volunteers were studied.
- ____ R_3 Purposive sampling from a specified population established the group studied.
- ____ R_4 Random selection from a specified population established the group studied.
- ____ R_5 The entire population was studied.

3 DATA QUALITY - TREATMENT

- ____ T_1 No theory; something undefined happened to the units studied.
- ____ T_2 No theory; treatment description incomplete, or detailed elsewhere.
- ____ T_3 No theory; treatment described in detail in the report.
- ____ T_4 Theory stated but no controls on variables.
- ____ T_5 Theory stated and mediating variables controlled.
- ____ T_6 Theory stated, mediating variables controlled, and techniques used to distribute possible extraneous variances.

4 DATA QUALITY - MEASUREMENT

- ____ M_1 Available information indicates instrument is <u>invalid</u> for this use.
- ____ M_2 Project Developed instrument with <u>low</u> validity (V), reliability (R), objectivity (O), or other instrument with <u>no info</u> about validity or data source.
- ____ M_3 Used Commercially Produced or Other-Project Developed instrument with <u>low</u> V,R,O for this application.
- ____ M_4 Used Project Developed instrument or Other-Project Developed instrument with <u>moderate</u> V, R, O for this application.
- ____ M_5 Used instrument which was Project Developed with <u>high</u> V, R, O or Other-Project developed with <u>high</u> V,R,O or Commercially Produced with <u>moderate</u> V, R, O for this application.
- ____ M_6 Used Commercially Produced instrument with <u>high</u> V, R, O for this application.

5 STATISTICAL ANALYSIS

- ____ A Appropriately analyzed
- ____ I Inappropriately analyzed
- ____ M Missing items - incomplete analysis